

Multivariate Descriptive Statistics

```
# Loading the dataset into a dataframe
df <- read_delim("../data/processed/wines.csv",
  ";",
  escape_double = FALSE,
  trim_ws = TRUE)
```

Analysis of the dataset

Quantitative attributes

```
## fixed_acidity    volatile_acidity    citric_acid    residual_sugar
## Min.      : 3.800    Min.      :0.0800    Min.      :0.0000    Min.      : 0.600
## 1st Qu.: 6.400    1st Qu.:0.2300    1st Qu.:0.2500    1st Qu.: 1.800
## Median : 7.000    Median :0.2900    Median :0.3100    Median : 3.000
## Mean      : 7.215    Mean      :0.3396    Mean      :0.3187    Mean      : 5.444
## 3rd Qu.: 7.700    3rd Qu.:0.4000    3rd Qu.:0.3900    3rd Qu.: 8.100
## Max.      :15.900    Max.      :1.5800    Max.      :1.6600    Max.      :65.800
## chlorides        free_sulfur_dioxide    total_sulfur_dioxide
## Min.      :0.00900    Min.      : 1.00      Min.      : 6.0
## 1st Qu.:0.03800    1st Qu.: 17.00      1st Qu.: 77.0
## Median :0.04700    Median : 29.00      Median :118.0
## Mean      :0.05602    Mean      : 30.52      Mean      :115.8
## 3rd Qu.:0.06500    3rd Qu.: 41.00      3rd Qu.:156.0
## Max.      :0.61100    Max.      :289.00      Max.      :440.0
## density          pH          sulphates          alcohol
## Min.      :0.9871    Min.      :2.720    Min.      :0.2200    Min.      : 8.00
## 1st Qu.:0.9923    1st Qu.:3.110    1st Qu.:0.4300    1st Qu.: 9.50
## Median :0.9949    Median :3.210    Median :0.5100    Median :10.30
## Mean      :0.9947    Mean      :3.219    Mean      :0.5313    Mean      :10.49
## 3rd Qu.:0.9970    3rd Qu.:3.320    3rd Qu.:0.6000    3rd Qu.:11.30
## Max.      :1.0390    Max.      :4.010    Max.      :2.0000    Max.      :14.90
## quality          type
## Min.      :3.000    Min.      :0.0000
## 1st Qu.:5.000    1st Qu.:0.0000
## Median :6.000    Median :0.0000
## Mean      :5.819    Mean      :0.2459
## 3rd Qu.:6.000    3rd Qu.:0.0000
## Max.      :9.000    Max.      :1.0000
```

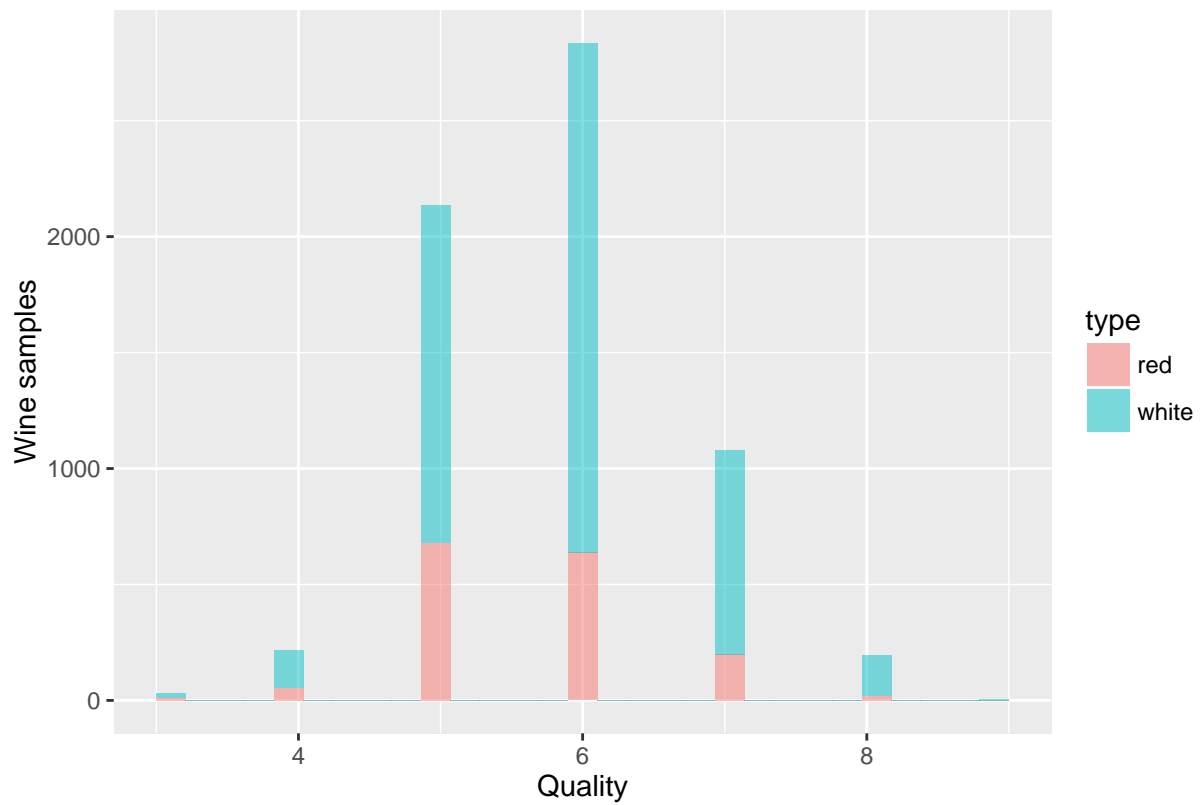
Quality histogram

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Table 1: Analysis of the attributes.

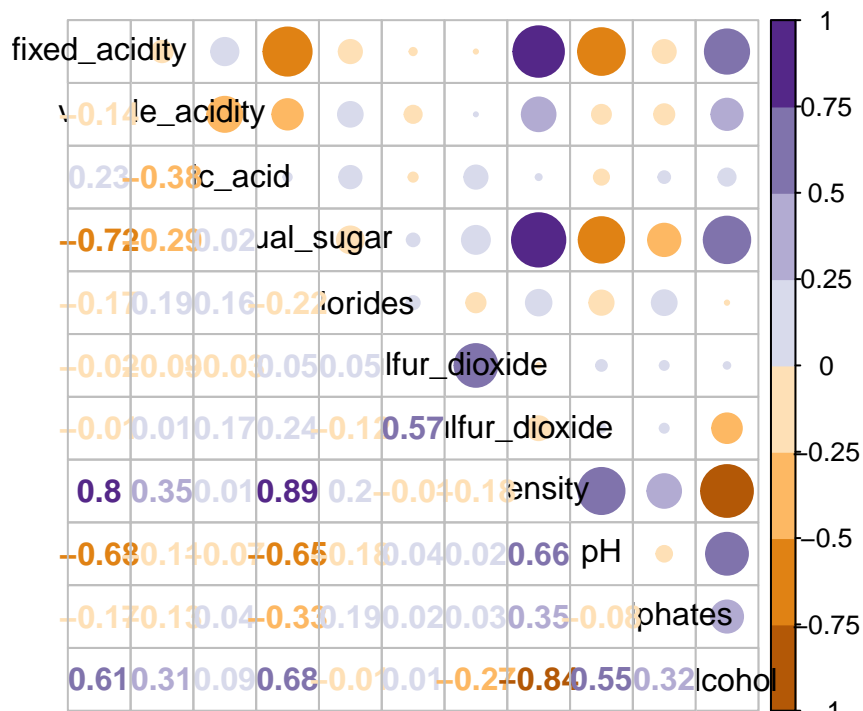
	min	max	mean	SD
fixed_acidity	3,800	15,900	7,215	1.296588
volatile_acidity	0,0800	1,5800	0,3396	0.164583
citric_acid	0,0	1,6600	0,3187	0.1452326
residual_sugar	0,600	65,800	5,444	4.758494
chlorides	0,00900	0,61100	0,05602	0.03503299
free_sulfur_dioxide	1,00	289,00	30,52	17.74849
total_sulfur_dioxide	6,0	440,0	115,8	56.52657
density	0,9871	1,0390	0,9947	0.002999095
pH	2,720	4,010	3,219	0.1608116
sulphates	0,2200	2,00	0,5313	0.148822
alcohol	8,00	14,90	10,49	1.192768

Histogram of wines quality: red vs. white



Correlation analysis

Partial correlations



Coefficients of determination and effective dependence coefficient of the R matrix

```
##      fixed_acidity      volatile_acidity      citric_acid
##      0.7957831        0.4866891        0.3776351
##      residual_sugar      chlorides      free_sulfur_dioxide
##      0.8616690        0.3870994        0.5324288
## total_sulfur_dioxide      density      pH
##      0.6598145        0.9371628        0.6058479
##      sulphates      alcohol
##      0.3513432        0.7908480
## [1] 0.4052199
```